Neural correlates of verbal joint action: ERPs reveal common perception and action systems in a shared-Stroop task

Keywords: Joint action, ERP, language, shared representations, Stroop

Şükrü Barış Demiral[1,2,*], Chiara Gambi[2], Mante S. Nieuwland[2], Martin J. Pickering[2]

[1]Laboratory of Neuroimaging, NIAAA, National Institutes of Health, Bethesda, MD, 20892, USA

[2]Department of Psychology, The University of Edinburgh, 7 George Square, Edinburgh, EH8 9JZ, UK

* Corresponding author:
Şükrü Barış Demiral, PhD
Laboratory of Neuroimaging (LNI)
NIAAA, NIH
10 Center Drive, Bldg 10
B2L124
Bethesda, MD, 20892
Phone: 301- 301.402.5861
e-mail: sukru.demiral@nih.gov

Abstract

Recent social-cognitive research suggests that the anticipation of co-actors' actions influences people's mental representations. However, the precise nature of such representations is still unclear. In this study we investigated verbal joint representations in a delayed Stroop paradigm, where each participant responded to one color after a short delay. Participants either performed the task as a single actor (single-action, Experiment 1), or they performed it together (joint-action, Experiment 2). We investigated effects of co-actors' action on the ERP components associated with perceptual conflict (Go N2) and response selection (P3b). Compared to single-action, joint-action reduced the N2 amplitude congruency effect when participants had to respond (Go trials), indicating that representing a co-actor's utterance helped to dissociate action codes and attenuated perceptual conflict for the responding participant. Yet, on NoGo trials the centro-parietal P3 (P3b) component amplitude increased for joint-action, suggesting that participants mapped the stimuli onto the co-actor's upcoming response as if it were their own response. We conclude that people represent others' utterances similarly to the way they represent their own utterances, and that shared perception-action codes for self and others can sometimes reduce, rather than enhance, perceptual conflict.

1. Introduction

Language use is a form of joint action (Clark, 1996). Whereas research has started to uncover the mechanisms that underlie non-linguistic forms of joint-action (Knoblich, Butterfill, & Sebanz, 2011), much less is known about the way people represent one another's utterances when using language together. There has been some recent interest in using event-related potential (ERP) to investigate linguistic joint action, and it appears that there is a tight coupling between brain activity in listeners and speakers (Kuhlen, Allefeld, Anders, & Haynes, 2015; Kuhlen, Allefeld, & Haynes, 2012). But there is less evidence about whether people represent their interlocutor's utterances as if it were their own.

To our knowledge, only one recent ERP study has contrasted the neural activation related to lexical processing versus the neural activation related to the prediction of another person's word in a joint task (Baus et al., 2014). In that study, a participant and a confederate took turns in naming pictures with high or low frequency names. Participants also engaged in lexical processing (as shown by a lexical frequency effect) when their partner was about to name them, suggesting that they were representing their partner's upcoming utterances.

However, it is unclear to what extent the findings from Baus et al. (2014) can be integrated with the dominant account of non-linguistic joint action, namely the co-representation account first proposed by Sebanz and colleagues (Sebanz, Knoblich, & Prinz, 2003; Sebanz, Knoblich, Prinz, & Wascher, 2006; Sebanz, Knoblich, & Prinz, 2005). According to this account, participants who are engaged in joint action represent their co-actor's task

and actions in the same format as their own task and actions. This account is supported by the joint spatial-compatibility effect observed in Go/NoGo versions of the Simon task (in which interference occurs when the spatial orientation of a stimulus is incompatible with a spatial feature of the response; see Simon,1990). In the joint task, one participant responded to stimuli of one color and their partner responded to stimuli of the other color, while both participants ignored spatial features of the stimuli (left or right orientation). Importantly, interference occurred in the joint version of this task, despite the fact that each participant performed either right or left responses, but not both. This is interpreted as evidence that the co-actor's response is represented in the same format as the actor's own response (Sebanz et al., 2003).

In an attempt to develop a more comprehensive understanding of the electrophysiological indices of verbal joint actions, the present ERP study used a delayed Stroop task. Crucially, participants performed the task either alone or taking turns with a co-actor. We asked whether participants represent their co-actor's upcoming utterances, and whether doing so leads to increased conflict, as predicted by the co-representation account.

1.1. ERP measures in joint-action research

Many aspects of the ERP signal can be informative about cognitive processing, even in the absence of a behavioral response. The ERP components commonly examined in joint-action research are the N2 and P3 components, because they are functionally associated with the underlying theoretical constructs of the experimental paradigms at hand (e.g., the Simon task). The N2 is a centro-frontal negative wave peaking around 150-250ms after stimulus onset. In paradigms where the experimental task inherently

measures perceptual and motor conflicts (e.g., Eriksen flanker tasks; Eriksen & Eriksen, 1974), the N2 is typically larger on incongruent Go trials than congruent Go trials (Heil, Osman, Wiegelmann, Rolke, & Hennighausen, 2000). In classical Stroop tasks, where participants respond on all trials, congruency-related frontal negativities emerge a bit later than the congruency effects mentioned above, mainly due to the relative complexity of the task (Liotti, Woldorff, Perez, & Mayberg, 2000). Based on such findings, the Go N2 is thought to index conflict detection (Donkers & van Boxtel, 2004). In addition, in Go/NoGo paradigms, the N2 is larger on NoGo trials than Go trials (Eimer, 1993), which has led others to suggest that the NoGo N2 indexes the non-motoric stage of inhibition in visual tasks (i.e., inhibition at the perceptual level before any motor actions are initiated) (Smith, Johnstone, & Barry, 2008).

The P3 is composed of a family of functionally and topographically distinct components. Relevant to our study, the frontocentral P3 (P3a) component emerges around 200-300ms, and occurs for the 'No-Go' items in oddball paradigms (Polich, 2007), generally following an N2 modulation. It is sometimes interpreted as an index of the completion of inhibitory processes (Tsai, Kuo, Jing, Hung, & Tzeng, 2006), or initial attentional component in cognitive control (Polich, 2007). In contrast, the centroparietal P3b component emerges between 200-600ms, and reflects mechanisms involved in contextual-update, uncertainty-resolution, and probability evaluation required for task performance (i.e., responding, mental calculation etc.) (Polich, 2007). Importantly, the P3b is related to stimulus-response mapping, a process that mediates and organizes the link between perceptual analysis and motor

preparation when these two are closely associated, such as planning and initiating a button press whenever you see one of two stimuli but not when you see the other stimulus (Kok, 2001; Verleger, Jaskowski, & Wascher, 2005). Thus, in contrast to the P3a, the P3b has been observed for the 'Go' items in oddball paradigms, making this component relevant for the stages of response representation and execution.

1.2. Representing another's upcoming actions

Sebanz et al. (2006) found that NoGo trials, on which the participant did not respond but the participant's partner did, showed larger frontal P3 (P3a) amplitude than NoGo trials on which neither the participant nor the participant's partner responded. This was taken as evidence of enhanced inhibition demands in joint tasks: Because participants represent their partner's actions as if they were their own on NoGo trials, they need to inhibit the tendency to respond overtly (Tsai, Kuo, Hung, & Tzeng, 2008; Tsai et al., 2006).

However, there are concerns about how to interpret responses on NoGo trials in these studies. First, while Tsai et al. (2006; 2008) found that the amplitude of the P3a was modulated by stimulus compatibility in the joint but not in the individual condition, Sebanz et al. (2006) found no effect of compatibility. Second, Tsai et al. (2006) found no difference between the amplitude of the NoGo N2 in the joint compared to the individual condition, even though this component is considered an index of inhibition. Moreover, Sebanz et al. (2006) and Tsai et al. (2008) did not report modulations of the NoGo N2 component.

In addition, the effect of representing another's potential response on one's own response preparation processes (i.e., on Go trials) is not clearly understood. In Sebanz et al. (2006) the amplitude of the P3b on Go trials was affected by compatibility in the individual condition but not in the joint condition. In contrast, Tsai et al. (2006) reported that on Go trials compatibility affected the amplitude of the P3b more in the joint condition than in the individual condition. Overall, these findings are problematic for the co-representation account. When it is their turn to act, co-actors should show increased interference from stimuli that evoke their partner's actions (as they do behaviourally; Knoblich et al., 2011), but it has proved difficult to identify ERP components that consistently show increased interference in joint compared to invidivual tasks.

In sum, although it appears that performing a task together with another participant affects ERP responses, the source of these effects is unclear, and so are their implications for theories of acting (and speaking) together. Most studies have assumed that the NoGo-related effects are due to inhibition, but a plausible interpretation of the range of P3a and N2 effects found across different joint-action studies is missing.

1.3. The present study

In this study, we provide a comprehensive test of the co-representation account by reporting analyses on both the N2 and the P3, for Go and NoGo trials. Importantly, we provide the first test of the account using a verbal task that induces interference. In two ERP experiments, we used a two-color delayed Stroop paradigm, in which each participant responded to only one color after a short delay. In the first experiment, participants performed the

task as single actors (single-action), so that their partner did not respond to the other color. Crucially, in the second experiment participants and partners performed the task together (joint-action), so that partners responded to the other color.

Based on previous studies, we expected neural responses to differ on NoGo trials (i.e., when participants do not have to respond) between the joint-action experiment and the single-action experiment. Such differential processing could reflect increased inhibition in the joint-action experiment. If so, we would expect larger N2 and larger P3a components (as indices of inhibition) for joint-action compared to single-action (Tsai et al., 2006). Alternatively, in the joint-action experiment, the participant could map the task-relevant property of the stimulus (i.e., the ink-color) to the co-actor's response, in the same way that she would map the stimulus to her own response on a Go trial. Specifically, we assume that the perceptual representation of the stimuli (e.g., word and color) and the responses participants make (e.g., utterances) share overlapping - or contrastive - features which may guide action selection. We will call this account *stimulus-response mapping* account. If this is the case, we would expect the amplitude of the NoGo P3b component (as an index of stimulus-response mapping) to be larger in the joint-action compared to the single-action experiment. Representing the co-actor's utterance will elicit responsiveness at the motor articulation stage in line with the perception-motor mediation account (Verleger et al., 2005).

In addition, following Tsai et al. (2006) and Sebanz et al. (2006), we predict the size of the congruency effect on Go trials to differ between the

single-action and the joint-action experiments. In the single-action experiment, participants will experience more conflict and hence greater N2 amplitude on incongruent than on congruent Go trials. If the co-representation account is correct, the congruency effect should be greater in the joint-action experiment than in the single-action experiment (following Tsai et al., 2006). Alternatively, if participants represent their partner's response as if it were their own, the co-actor's response would become part of the participant's own response set and may induce less conflict for the future response. If this is the case, the mental operations relevant for alerting and conflict detection could be moderated in the joint-action experiment compared to the single-action experiment, leading to decreased perceptual conflict (as indexed by the Go N2 component).

Note that the N2 component has also been investigated in other linguistic domains: comparing bilinguals with monolinguals (Fernandez, Tartar, Padron, & Acosta, 2013), and in bilingual tasks where subjects switch between L1 and L2 while naming objects (Liu, Rossi, Zhou, & Chen, 2014; Verhoef, Roelofs, & Chwilla, 2009). Generally, studies using linguistic picture naming tasks elicited an N2 in a later time window (around 300-360ms) than studies using more standard tasks such as tone detection (240-300ms). This might be due to the longer time needed for more complex perceptual and semantic operations (i.e., extracting object information from memory), or due to linguistic operations during naming. In our task, we tested for the presence of an N2 effect in both an earlier and a later time window.

In sum, if the stimulus-response mapping account is correct, we expect that: (1) participants will represent their partner's upcoming utterance on NoGo trials, as indexed by a larger P3b amplitude on NoGo trials in the joint-

action compared to the single-action context; and (2) participants will represent their partner's potential utterance on incongruent Go trials, with this leading to decreased perceptual conflict, as indexed by a reduced congruency effect on the amplitude on the N2 component on Go trials in the joint-action compared to the single-action context. Alternatively, if the co-representation account is correct, inhibition demands should be larger in the joint-action compared to the single-action context, and this should be reflected in larger N2 and P3a on joint- than single-action NoGo trials. Moreover, if the congruency effect in joint-action Go trials exhibits a larger N2 than single-action Go trials, then this might also be taken as evidence favoring the co-representation account due to increased perceptual conflict in social contexts.

2. Results

2.1. Behavioral Results:

There were some time-outs but the overall count was extremely low (<1%), perhaps because our task involved only two colors and/or because of the response delay. Thus, we included all the trials in our analysis. Feedback response times of the actor in the Congruent versus Incongruent NoGo trials were very similar: mean(std), 934ms(167ms), and 930ms(164ms) respectively.

2.2. ERP Results:

2.2.1. Go Trials:

ERP waves generated on Go trials can be seen in **Figure 1**

<INSERT Figure 1 HERE>

N2 window(a) 180-240ms:

There was a significant interaction of Experiment x Congruity, $F(1,96)=3.10$, $p<.05$. Incongruent trials elicited larger (more negative) N2 components than congruent trials in the single-action experiment, $F(1,24)=5.08$, $p<.05$, whereas no such difference was found for the joint-action experiment, $F<1$ (**Figure 2**).

<INSERT Figure 2 HERE>

N2 window(b) 220-350ms:

There was a main effect of Congruity, $F(1,48)=5.63$, $p<.05$. In addition, there was a significant interaction of Experiment x Congruity, $F(1,96)=5.77$, $p<.05$. In only the single-action experiment, incongruent trials elicited an N2 effect, $F(1,24)=7.49$, $p<.05$. For the joint-action experiment, there was no effect of Congruity $F<1$.(**Figure 3**).

<INSERT Figure 3 HERE>

P3b window: 320-800ms

There was no main effect of Congruity and no main effect of Experiment, both $Fs<1$. The interaction of Experiment x Congruity was significant, $F(1,96)=5.50$, $p<.05$. However, resolving this interaction showed that there was no reliable effect of congruency in either the joint-action, $F(1,24)=2.85$, $p=.10$, or the single-action experiment, $F(1,24)=2.09$, $p=.16$. Similarly, the two groups of

participants did not differ in the amplitude of the P3b component on either congruent or incongruent trials, both Fs<1.

Early P3b window: 320-420ms

There was neither a main effect nor an interaction for this time window.

Late P3b window: 420-800ms

The interaction of Experiment x Congruity was significant, $F(1,96)=4.52$, $p<.05$), but neither single-action nor joint-action experiment showed a significant effect of congruity. Similarly, when resolved with Congruity, no effect of Experiment was observed.

2.2.2. NoGo Trials:

ERP waves generated on NoGo trials can be seen in **Figure 4**.

<INSERT Figure 4 HERE>

N2 window(a) 180-240ms:

There was no main effect of Congruity, $F<1$, and no main effect of Experiment, $F(1,48)=1.46$, $p=.23$. The interaction of Experiment x Congruity was also not significant, $F<1$.

N2 window(b) 220-350ms: There was a Region x Congruity interaction $F(2,96)=5.21$, $p<.01$), but resolving it for Region did not reveal any effect of Congruity for any region ($F<1$).

P3b window: 320-800ms

P3b amplitude was significantly larger in the Joint compared to the Single experiment in the posterior electrodes, as indicated by a significant Experiment x Region interaction, $F(2,96)=11.31$, $p<.001$. Only Posterior electrodes showed a significant effect of Experiment, $F(1,48)=7.16$, $p<.01$, whereas Anterior electrodes and Central electrodes did not show an effect of Experiment, both Fs <1. The interaction between Experiment and Topography was also significant, $F(2,96)=3.33$, $p<.039$. Resolving this effect showed that Experiment was significant in the left hemisphere electrodes, $F(1,48)=4.1$, $p<.05$, but not in the right hemisphere, or in the middle electrodes, both F's <1. No other effect or interaction reached significance (**Figure 5).**

<INSERT Figure 5 HERE>

Early P3b window: 320-420ms

Similar to the large time window, P3b amplitude was significantly larger in the Joint compared to the Single experiment in the posterior electrodes, as indicated by a significant Experiment x Region interaction, $F(2,96)=3.31$, $p<.05$. Only Posterior electrodes showed a significant effect of experiment, $F(1,48)=4.72$, $p<.05$. (**Figure 6**).

<INSERT Figure 6 HERE>

Late P3b window: 420-800ms:

Experiment interacted with Topography, $F(2,96)=4.17$, $p<.05$, and with Region, $F(2,96)=13.31$, $p<.05$. While none of the topographies showed a main

effect of Experiment ($F<1$), P3b amplitude was larger on the posterior region electrodes, $F(1,48)=7.02$, $p<.05$. In addition, there was a 3-way interaction between Experiment, Region and Topography, $F(4,192) =2.95$, $p<.05$. Resolving this interaction for Region, there was a Topography x Experiment interaction in the frontal regions, $F(2,96)=7.88$, $p<.001$, but Experiment was not found to be significant in any of the topographies in the omnibus Anova. When this 3-way interaction was resolved with Topography first, all the topographies showed a significant Experiment x Region interaction ($p<.05$). Resolving this interaction further for each topography showed that only in the posterior region there was a significant effect of Experiment in every topography ($p<.05$) (**Figure 7**). Therefore, we found no evidence that the amplitude of the anterior P3a differed between the experiments.

<INSERT Figure 7 HERE>

3. Discussion

On Go trials in the Single-action experiment, we found an N2 effect reflecting perceptual conflict on Incongruent relative to Congruent trials. This effect was not present in the Joint–action experiment. Therefore, when participants took turns naming with their partner, they experienced reduced perceptual conflict from the printed word on incongruent trials. We suggest this occurred because performing a joint verbal task leads to the activation of the co-actor's potential utterance (which in our task was the same as the interfering printed word), so that this utterance is no longer represented as a conflicting utterance. Thus, we interpret this as evidence that people form

common perceptual representations of their own and a co-actor's utterances. Following the N2, on Go trials we observed a P3 component that was particularly pronounced at posterior sites (P3b, see Figure 1), and whose amplitude did not differ between the joint and the single-action experiment, nor between congruent and incongruent trials. It is possible that this component reflects post-perceptual response execution operations that are insensitive to perceptual conflicts.

On NoGo trials, we found a larger P3b amplitude in the Joint-action experiment than the Single-action experiment. In line with the stimulus-response mapping account and the claim that the P3b reflects the mediation between perceptual analysis and response initiation (Kok, 2001; Verleger et al., 2005), this finding suggests that stimuli triggering the co-actor's response elicited a perceptual-motor mapping in the participant. This further supports the idea that people represent others' utterances as if they were their own utterances when those utterances are produced as part of a joint action (and participants are required to monitor them). However, this finding is open to alternative interpretations, as we discuss in the next section.

Contrary to the predictions of co-representation account, on the NoGo trials, we did not find a larger fronto-central P3 (P3a) in the Joint- than the Single-action experiment. Previous studies (e.g., Sebanz et al., 2006) have interpreted such a difference as evidence that inhibition demands are greater in joint than single tasks. Similarly, the N2 did not differ in the Joint- and the Single-action experiment (as in Tsai et al., 2006). Also, the N2 generally occurs together with the P3a on NoGo trials (Folstein & Van Petten, 2008; Patel & Azzam, 2005), and should therefore be larger for joint than individual

conditions, if the co-representation account is correct. This provides no evidence that participants exert stronger inhibition in joint tasks in order to avoid responding when it is their partner's turn. Note that some linguistic experiments (e.g., using picture naming; (Liu et al., 2014; Verhoef et al., 2009) have found a later inhibition-related N2 component (around 300-360ms), perhaps reflecting more complex perceptual and semantic processing, but we did not find any sign that the component was delayed.

Differences between the results of our experiments and those of shared Simon experiments might reflect differences between Simon and Stroop tasks – though we also note that Sebanz et al. (2006) and Tsai et al. (2006) found results that were not fully compatible with each other. One difference relates to arbitrariness of the stimulus-response relationship. In the shared Simon task, the relationship between stimulus (color) and response (Go or NoGo) is arbitrary (though the direction of hand-pointing of course has a non-arbitrary relation to response). In the shared Stroop task, the relationship between stimulus (color) and response (name or not name) is arbitrary in one sense (that the participant responds to one color but not the other) but is non-arbitrary in another sense (that the participant names the color appropriately). It may be that differential effects of inhibition may reflect the cognitive differences between spatial and non-spatial coding.

3.1 Joint actions, attention, and cognitive load

Taking others into account requires the allocation of attentional networks. Recent findings suggested that joint-action representations in social Simon tasks could be driven by attention-capturing factors, given that acting together with inanimate objects can have the same effect as acting together

with real co-participants provided those objects can capture participants' attention (Dolk, Hommel, Prinz, & Liepelt, 2013; Doneva & Cole, 2014).

In a set of recent theoretical approaches, namely Referential Coding and Theory of Event Coding (Dolk et al., 2014; Hommel, Musseler, Aschersleben, & Prinz, 2001), joint effects are attributed to the overlapping perceptual (stimulus) and motor (action) codes provided in the environment. These theories were mainly proposed to explain both social and non-social Simon effects, and they claim that contextual cues may influence the action representations due to the conflicting codes they provide, and this causes action selection difficulty. For instance, participants may experience difficulty when the action codes they follow as a default (e.g., press the "right" button), and the perceptual codes provided in the environment – together with their representational consequences - overlap (e.g., showing a ticking clock on their "right side"). These effects can be elicited purely via attentional mechanisms. For instance, external stimuli presented in the same locations might use up attentional sources merely because they are presented in the same locations as the action direction. Thus, attentional mechanisms might play a role in the allocation of mental resources in joint action tasks.

In our study, we asked ERP participants to provide feedback to their co-actors about their performance in the joint-action experiment but not in the single-action experiment. This requirement might have affected the allocation of attention, and raises the possibility that the observed enhancement of the P3b on joint NoGo compared to Single NoGo trials partly resulted from increased attention to NoGo stimuli in the joint-action experiment. We think this is unlikely, though, given that the P3a, not the P3b, is usually considered

as an index of attention allocation (Polich, 2007). Actually, this interpretation of the P3a provides a potential link between the findings of Sebanz et al. and the findings of Dolk et al., specifically that the P3a observed in the joint Simon task might be a by-product of attentional switching in space that may be caused by spatial cues. In our study, the requirement of providing feedback to their co-actor could have encouraged participants to process their stimuli as if they were their own, and map those stimuli to the expected response, to facilitate the monitoring task. Attending to a co-actor's utterances, gestures, and other relevant cues is of course necessary for successful joint action. An interesting question is whether similar findings might occur if participants perform a monitoring task on NoGo trials in a single action context.

Joint-action and single-action in our study may have differed in cognitive load as well. Monitoring partner's performance was a secondary task for participants in the joint-action experiment, hence making this task cognitively more demanding. The ERP literature on cognitive load and resource allocation shows that expectancy about the task difficulty and performing a task along with a secondary task (i.e., dual-task performance) have dissociable effects on ERP components, particularly on the P3b component. For instance, Wilson et al. showed that stimulus warning about the difficulty of the upcoming performance elicited a larger P3b component for harder tasks compared to easier tasks (Wilson, Swain, & Ullsperger, 1998). In addition, Kok reported that greater task difficulty elicited a smaller P3b amplitude, but a reverse pattern emerged when the primary task was accompanied by a secondary task (Kok, 1997). Note that, contrary to our study, the secondary tasks used in these experiments were not social in

nature (individuals performed the task alone), and they require responses to task-irrelevant stimuli (e.g., responding to deviant tones in an odd-ball paradigm). Additionally, these studies contrasted different levels of cognitive load (i.e., task difficulty) containing memory-dependent features (e.g., number of digits to memorize) of the primary task stimuli.

Nevertheless, it is worth considering how increased cognitive load in the joint-action experiment could have affected our findings. Interestingly, Baus et al. (2014) interpreted some of their findings from their joint linguistic task interms of increased load. When participants were naming pictures on Go trials in the joint version of the task, they found a frequency effect on P3b amplitude, but there was no frequency effect on Go trials in the individual condition. They attributed this difference to greater cognitive demands in the joint condition (rather than to the effects of joint representations; see p. 405). If we take the frequency effect as a measure of task difficulty, Baus et al.'s results are indeed compatible with Kok's (1997) finding that P3b amplitude increases for more difficult tasks under greater cognitive load.

Similarly, in our study, if we treat congruity as an index of task difficulty, and assume that monitoring one's performance increased cognitive load (similarly to performing a secondary task), we should expect an increased congruity effect on P3b amplitude in the joint Go compared to single Go trials. While we found that P3b amplitude was higher in joint trials than the single trials in the Go condition overall, we did not observe any effect of congruity on these measures. It is still possible that the overall joint performance effect on P3b is due to the increased task demands as proposed above. Future studies could address this issue directly, and dissociate dual

linguistic task performances in single versus social contexts with varying task demands, to evaluate their influence on representing others' action plans.

4. Conclusion

We showed that people taking part in a joint verbal task come to form perceptual and motor representations of their co-actor's upcoming utterances that are similar to the representations underlying the preparation of their own utterances. This finding is not only relevant for theories of acting together in general (as discussed above), but also for understanding the cognitive underpinnings of joint language use in particular. Clearly, a joint Stroop task is far removed from everyday conversation. Importantly, we do not claim that the effects we have reported are uniquely linguistic in nature. To the contrary, they are likely to be grounded in basic perception and action mechanisms, which are relevant to verbal as to other types of actions. However, the task mimicked at least one key aspect of dialogue, namely the need to closely monitor both one's own and one's partner's utterances. The activation of the co-actor's utterance is compatible with theories of language use in which speakers represent their own and their addressees' potential utterances in a common format (Pickering & Garrod, 2013).

We have showed that perceptual conflict is reduced in a joint Stroop task compared to an individual Stroop task. On the basis of this finding, we have proposed that people represent their partner's utterances within the perceptual-motor action system as if they were their own utterances. We suggest that this constitutes one basic cognitive mechanism that facilitates joint action.

5. Methods and Materials

In Experiment 1, two participants sat side-by-side. We recorded ERPs for participant A but not B. Participant A was presented with iterations of stimuli: the word *red* in red or the word *green* in green (i.e., congruent stimuli); and the word *red* in green or the word *green* in red (i.e., incongruent stimuli). Participant A responded to one color (balanced across participants) in a Go-NoGo design, but waited for a signal 1200ms after stimulus onset before responding. If participant A responded correctly, then Participant B was to respond "YES"; if participant A responded incorrectly, then Participant B was respond "NO" and press a button to indicate a wrong response (that was also marked in the EEG signal). Participants' utterances were also recorded via E-Prime response device using a microphone.

In Experiment 2, Participant A and B both responded to the stimuli, with Participant A responding to one color and Participant B responding to the other color. Again, we recorded ERPs for participant A but not B. In addition, both participants provided feedback to their partners' responses. That is, in Experiment 2, Participant A also had to monitor the co-actor's actions.

Note that the version of the Stroop task we used had three major differences from the traditional Stroop designs used in previous ERP studies. First, only two colors were used, which makes this task easier than the standard Stroop task with three or more colors (MacLeod, 1991; Peterson et al., 2002). Second, we employed a Go/NoGo version of the task (name vs. not-name), so that each participant responded to only one of the colors rather than naming both colors. Third, we used a delayed response paradigm, where participants did not respond immediately. We did so to (a) reduce highly variable and temporally unstable speech-related muscle artifacts, (b)

dissociate motor preparation and motor response stages as much as possible, and (c) provide stable speech onset times and prevent any confounds in the ERP measures due to speech onset delays (Strijkers, Holcomb, & Costa, 2012). In addition, unlike Baus et al. (2014), we used different participants in the single and joint version of the task, and they were paired with other participants randomly rather than with confederates. We thus prevented any potential effect of task-order and confederate bias on behavior and ERPs in the joint experiment.

5.1. Experiment 1: Single action

5.1.1. Participants

Fifty students from the University of Edinburgh, aged 18 to 25, were paid £6 per hour to participate. All the participants were native speakers of English. ERPs were recorded from 25 participants (M_{age} = 22.2, SD_{age} = 3.8; right-handed; 15 female), and the other 25 participants were randomly paired with the ERP participants (see below). All participants reported normal or corrected-to-normal vision and no neurological or psychiatric problems.

5.1.2. Stimuli

There were four stimuli: the word *red* in red, the word *green* in green (i.e., congruent stimuli); and the word *red* in green, the word *green* in red (i.e., incongruent stimuli).

5.1.3. Procedure

Participants completed 360 trials (90 trials per condition) in 3 equal blocks. In each session, one ERP participant and one control participant sat side by side in front of a single computer screen (90cm away from the

participants). The stimuli were presented using E-prime software (Psychology Software Tools Inc., Pittsburgh, PA). The stimulus word appeared for 500ms, followed by a blank screen for 700ms, followed by a cross for 2000ms. The ERP participant was instructed to name only one color (either RED or GREEN) as soon as the cross appeared. ERP participants were told to minimize eye-movements and blinks during each trial. The control participant was instructed to respond "Yes" when the ERP participant responded correctly and "No" otherwise. The control participant also pressed a response button whenever the ERP participant made an error. The responses and onset latencies of both participants were recorded with the E-Prime Response Box. We monitored both response utterances and the button presses for the detection of incorrect responses. The experiment took about two hours, including cap and electrode preparation.

5.1.4. EEG recording

EEG activity was recorded by a BioSemi ActiveTwo system (http://www.biosemi.com), using 64 EEG channels configured according to the 10-20 system. Two additional electrodes were placed on the left and right mastoids, and 4 additional electrodes recorded the EOG (horizontal cantus of the left and right eye for HEOG, above and below the right eye for VEOG). The EEG sampling rate was 512Hz.

5.1.5. EEG signal processing

We used a combination of in-house algorithms and EEGLAB software (Makeig, Debener, Onton, & Delorme, 2004). After DC-detrending, a FIR filter was used for high- and low-pass filtering the data with half-amplitude cut-off values of 0.1Hz and 20Hz respectively (12dB/octave). The data were then re-

referenced to the mean of the mastoid electrodes. Epochs were selected from -200ms to 1200ms relative to the onset of the stimulus word. Ocular artefacts were automatically corrected with independent component analysis (ICA) in the ADJUST toolbox (Mognon, Jovicich, Bruzzone, & Buiatti, 2010). This procedure requires gross artefact correction before applying ICA correction, which resulted in the rejection of trials for each condition (Incongruent Go/Congruent Go/Incongruent NoGo/Congruent NoGo) with mean(SD) as follows: 1.84(1.99), 1.81(2.04), 1.92(2.13), and 1.76(1.64) for joint performance, and 4.41(6.8), 4.88(8.92), 3.88(6.98), and 4.08(8.4) for the single performance. Epochs were then baselined to the average value between -100ms and 0ms relative to stimulus onset.

5.1.6. EEG electrode regions of interest

The ERP analysis followed Sebanz et al. (2006), using nine electrodes in three Regions: anterior (F3/Fz/F4), central (C3/Cz/C4), posterior (P3/Pz/P4), and in three different Topographic fields: left (F3/C3/P3), center (Fz/Cz/Pz), and right (F4/C4/P4). In order to be able to compare our results with the existing literature from joint performance literature (Sebanz et al. 2006) and tasks which are linguistic in nature involving task switching (Liu et al., 2014), we defined two different time-windows of interest for N2 and P3b components.

5.1.7. Statistical analysis of the ERP data

Following Sebanz et al. (2006), we conducted repeated-measures analysis of variance (ANOVA) for Go and NoGo trials separately. As mentioned in Sebanz et al. (2006), Go and NoGo trials represent two distinct processing routines, namely “action planning” and “action control”, which are

the two main ingredients of the theories of joint action. In line with these authors, we also assumed that analysing these two trial types separately would provide a better understanding of the action and inhibition effects that differ between single and social behavior, and would provide statistical simplicity. For each of these analyses, we introduced Region (Anterior/Central/Posterior), Topography (Left/Center/Right), and Congruity (Congruent/Incongruent) as within-subject variables, and Experiment (Single versus Joint) as a between-subject variable. A Huynh-Feldt correction was applied when evaluating effects with more than one degree of freedom. A Bonferroni correction was used for multiple comparisons.

5.2. Experiment 2: Joint action

5.2.1. Participants

Fifty further participants from the same population as Experiment 1 took part under the same terms (M_{age}= 22.9, SD_{age} =3.9; right-handed; 15 female). Participants were randomly paired as in Experiment 1.

5.2.2. Procedure

The procedure was identical to Experiment 1, except (a) the Control participant responded to the opposite color from the ERP participant, and (b) both participants gave feedback about each other's performance by uttering "Yes" or "No" after their main responses and pressing a button whenever their partner made an error.

Acknowledgements

This work was supported by funding provided by the Department of Psychology, The University of Edinburgh.

References

Baus, C., Sebanz, N., Fuente, V. D., Branzi, F. M., Martin, C., & Costa, A. (2014). On predicting others' words: Electrophysiological evidence of prediction in speech production. *Cognition, 133*(2), 395-407. doi: 10.1016/j.cognition.2014.07.006

Clark, H. H. (1996). *Using language*. Cambridge England ; New York: Cambridge University Press.

Dolk, T., Hommel, B., Colzato, L. S., Schutz-Bosbach, S., Prinz, W., & Liepelt, R. (2014). The joint Simon effect: a review and theoretical integration. *Front Psychol, 5*, 974. doi: 10.3389/fpsyg.2014.00974

Dolk, T., Hommel, B., Prinz, W., & Liepelt, R. (2013). The (not so) social Simon effect: a referential coding account. *Journal of Experimental Psychology-Human Perception and Performance, 39*(5), 1248-1260. doi: 10.1037/a0031031

Doneva, S. P., & Cole, G. G. (2014). The role of attention in a joint-action effect. *PLoS One, 9*(3). doi: 10.1371/journal.pone.0091336

Donkers, F. C. L., & van Boxtel, G. J. M. (2004). The N2 in go/no-go tasks reflects conflict monitoring not response inhibition. *Brain and Cognition, 56*(2), 165-176. doi: 10.1016/j.bandc.2004.04.005

Eimer, M. (1993). Effects of attention and stimulus probability on ERPs in a Go/Nogo task. *Biol Psychol, 35*(2), 123-138.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. . *Perception & Psychophysics, 16*, 143-149.

Fernandez, M., Tartar, J. L., Padron, D., & Acosta, J. (2013). Neurophysiological marker of inhibition distinguishes language groups on a non-linguistic executive function test. *Brain and Cognition, 83*(3), 330-336. doi: DOI 10.1016/j.bandc.2013.09.010

Folstein, J. R., & Van Petten, C. (2008). Influence of cognitive control and mismatch on the N2 component of the ERP: a review. *Psychophysiology, 45*(1), 152-170. doi: 10.1111/j.1469-8986.2007.00602.x

Heil, M., Osman, A., Wiegelmann, J., Rolke, B., & Hennighausen, E. (2000). N200 in the Eriksen-task: Inhibitory executive processes? *Journal of Psychophysiology, 14*(4), 218-225. doi: 10.1027//0269-8803.14.4.218

Hommel, B., Musseler, J., Aschersleben, G., & Prinz, W. (2001). The Theory of Event Coding (TEC): a framework for perception and action planning. *Behav Brain Sci, 24*(5), 849-878; discussion 878-937.

Knoblich, G., Butterfill, S., & Sebanz, N. (2011). Psychological research on joint action: Theory and data. *Psychology of Learning and Motivation: Advances in Research and Theory, 54*, 59-101. doi: 10.1016/B978-0-12-385527-5.00003-6

Kok, A. (1997). Event-related-potential (ERP) reflections of mental resources: a review and synthesis. *Biol Psychol, 45*(1-3), 19-56.

Kok, A. (2001). On the utility of P3 amplitude as a measure of processing capacity. *Psychophysiology, 38*(3), 557-577. doi: 10.1017/S0048577201990559

Kuhlen, A. K., Allefeld, C., Anders, S., & Haynes, J. D. (2015). Towards a Multi-Brain Perspective on Communication in Dialogue. In R. M. Willems (Ed.), *Cognitive Neuroscience of Natural Languge Use*. Cambridge: Cambridge University Press.

Kuhlen, A. K., Allefeld, C., & Haynes, J. D. (2012). Content-specific coordination of listeners' to speakers' EEG during communication. *Frontiers in Human Neuroscience, 6*. doi: 10.3389/Fnhum.2012.00266

Liotti, M., Woldorff, M. G., Perez, R., & Mayberg, H. S. (2000). An ERP study of the temporal course of the Stroop color-word interference effect. *Neuropsychologia, 38*(5), 701-711. doi: Doi 10.1016/S0028-3932(99)00106-2

Liu, H., Rossi, S., Zhou, H., & Chen, B. (2014). Electrophysiological evidence for domain-general inhibitory control during bilingual language switching. *PLoS One, 9*(10), e110887. doi: 10.1371/journal.pone.0110887

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: an integrative review. *Psychol Bull, 109*(2), 163-203.

Makeig, S., Debener, S., Onton, J., & Delorme, A. (2004). Mining event-related brain dynamics. *Trends in cognitive sciences, 8*(5), 204-210. doi: 10.1016/j.tics.2004.03.008

Mognon, A., Jovicich, J., Bruzzone, L., & Buiatti, M. (2010). ADJUST: An automatic EEG artifact detector based on the joint use of spatial and temporal features. *Psychophysiology*. doi: 10.1111/j.1469-8986.2010.01061.x

Patel, S. H., & Azzam, P. N. (2005). Characterization of N200 and P300: selected studies of the Event-Related Potential. *Int J Med Sci, 2*(4), 147-154.

Peterson, B. S., Kane, M. J., Alexander, G. M., Lacadie, C., Skudlarski, P., Leung, H. C., . . . Gore, J. C. (2002). An event-related functional MRI study comparing interference effects in the Simon and Stroop tasks. *Brain Res Cogn Brain Res, 13*(3), 427-440.

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behav Brain Sci, 36*(4), 329-347. doi: 10.1017/S0140525X12001495

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clin Neurophysiol, 118*(10), 2128-2148. doi: 10.1016/j.clinph.2007.04.019

Sebanz, N., Knoblich, G., & Prinz, W. (2003). Representing others' actions: Just like one's own? *Cognition, 88*(3), B11-B21.

Sebanz, N., Knoblich, G., Prinz, W., & Wascher, E. (2006). Twin peaks: an ERP study of action planning and control in co-acting individuals. *Journal of Cognitive Neuroscience, 18*(5), 859-870. doi: 10.1162/jocn.2006.18.5.859

Sebanz, N., Knoblich, G., & Prinz, W. G. (2005). How two share a task: Corepresenting stimulus-response mappings. *Journal of Experimental Psychology-Human Perception and Performance, 31*(6), 1234-1246. doi: Doi 10.1037/0096-1523.31.6.1234

Simon, J. R. (1990). The effects of an irrelevant directional cue on human information processing. In R. W. Proctor & T. G. Reeve (Eds.), *Advances in psychology* (Vol. Vol. 65.): Oxford, England: North-Holland.

Smith, J. L., Johnstone, S. J., & Barry, R. J. (2008). Movement-related potentials in the Go/NoGo task: the P3 reflects both cognitive and motor inhibition. *Clin Neurophysiol, 119*(3), 704-714. doi: 10.1016/j.clinph.2007.11.042

Strijkers, K., Holcomb, P. J., & Costa, A. (2012). Conscious intention to speak proactively facilitates lexical access during overt object naming. *J Mem Lang, 65*(4), 345-362. doi: 10.1016/j.jml.2011.06.002

Tsai, C. C., Kuo, W. J., Hung, D. L., & Tzeng, O. J. (2008). Action co-representation is tuned to other humans. *Journal of Cognitive Neuroscience, 20*(11), 2015-2024. doi: 10.1162/jocn.2008.20144

Tsai, C. C., Kuo, W. J., Jing, J. T., Hung, D. L., & Tzeng, O. J. L. (2006). A common coding framework in self-other interaction: evidence from joint action task. *Experimental Brain Research, 175*(2), 353-362. doi: DOI 10.1007/s00221-006-0557-9

Verhoef, K., Roelofs, A., & Chwilla, D. J. (2009). Role of inhibition in language switching: evidence from event-related brain potentials in overt picture naming. *Cognition, 110*(1), 84-99. doi: 10.1016/j.cognition.2008.10.013

Verleger, R., Jaskowski, P., & Wascher, E. (2005). Evidence for an integrative role of P3b in linking reaction to perception. *Journal of Psychophysiology, 19*(3), 165-181. doi: 10.1027/0269-8803.19.3.165

Wilson, G. F., Swain, C. R., & Ullsperger, P. (1998). ERP components elicited in response to warning stimuli: the influence of task difficulty. *Biol Psychol, 47*(2), 137-158.

Figure Legends

Figure 1. ERP waveforms on Go trials. Centro-frontal N2 peaks are marked by arrows. First peak is in the 180-240ms, and the later peak is in the 220-350ms window. Negative polarity is plotted up, and positive polarity is plotted down.

Figure 2. Topographical view of the first N2 component in 180-240ms time window. Top-left scale is for the images A1, A2, B1, and B2. Top-right scale is for the images A3 and B3. Bottom-right scale is for the C image.

Figure 3. Topographical view of second N2 component in 220-350ms time window. Top-left scale is for the images A1, A2, B1, and B2. Top-right scale is for the images A3 and B3. Bottom-right scale is for the C image.

Figure 4. ERP waveforms on NoGo trials. Parietal P3b peak and the following sustained wave are marked by arrows. Negative polarity is plotted up, and positive polarity is plotted down.

Figure 5. Topographical view of the parietal P3b component in 320-800ms time window. Top-left scale is for the images A1, A2, B1, and B2. Top-

right scale is for the images A3 and B3. Bottom-right scale is for the C image. Go conditions (B-C) are given for visual comparison purposes.

Figure 6. Topographical view of the peak segment of the P3b in 320-420ms time window. Top-left scale is for the images A1, A2, B1, and B2. Top-right scale is for the images A3 and B3. Bottom-right scale is for the C image. Go conditions (B-C) are given for visual comparison purposes.

Figure 7. Topographical view of the later segment of the P3b wave in 420-800ms time window. Top-left scale is for the images A1, A2, B1, and B2. Top-right scale is for the images A3 and B3. Bottom-right scale is for the C image. Go conditions (B-C) are given for visual comparison purposes.